Observations on Determinization of Büchi Automata

Christoph Schulte Althoff, Wolfgang Thomas, and Nico Wallmeier

RWTH Aachen, Lehrstuhl Informatik 7, 52056 Aachen, Germany
{althoff, thomas, wallmeier}@i7.informatik.rwth-aachen.de

Abstract. The two determinization procedures of Safra and Muller-Schupp for Büchi automata are compared, based on an implementation in a program called OmegaDet.

1 Introduction

A central result in the theory of ω-automata is McNaughton's Theorem [1]. In its original formulation it says that a nondeterministic Büchi automaton can be converted into a deterministic Muller automaton. Many constructions have been proposed to show this determinization result (cf. [2], [3] for hints to the literature). In most cases, the target automaton is a deterministic Rabin automaton, which can be considered as a special form of Muller automaton. It is well-known that the blow-up in number of states has to be greater than in the classical subset construction; there is a lower bound of $2^{O(n \log n)}$ for the number of states of a deterministic Rabin automaton, given a Büchi automaton with n states ([4], [5]).

Safra [6] was the first to find a construction which matches this lower bound. It seems now the standard way of showing Büchi automata determinization. But there is a second construction, again matching the lower bound, due to Muller and Schupp [7]. Teaching experience of the second author indicates that the Muller-Schupp proof can be explained more easily (the reader can judge him/herself below). The two constructions are sufficiently different to justify a closer comparison. This is the aim of present paper.

The study is based on an implementation of the two algorithms in a program called OmegaDet. As it turns out, such an implementation is necessary not only for a serious experimental performance comparison of the two procedures but also for a better conceptual understanding of their behaviour. The two algorithms are too involved to be analyzed by hand if one is interested in studying say a dozen examples. We report here on some insights we obtained in this investigation, both regarding a better understanding of the characteristics (and the similarities) of two the algorithms and of their performance. We observe that (apart from some peripheral cases), the Safra algorithm uses stronger abstractions than the one of Muller-Schupp and thus yields smaller automata. (Maybe this is a reason for the difficulties the Safra algorithm poses for exposition in lectures.) Moreover, our experiments led to an improvement of the Muller-Schupp procedure which can reduce the state space of the target automaton.

The paper is structured as follows. In the subsequent Section 2 we present the determinization procedures, with an emphasis on singling out those points where they coincide and where they differ. In this exposition, we use the insights from our experiments; we start with an explanation of the Muller-Schupp algorithm and more briefly discuss the Safra construction. We do not give any details about the correctness proofs. In Section 3 we give a brief introduction to (the user's view of) the program OmegaDet, in which also the subset based construction of Hayashi and Miyano [8] for co-Büchi automata determinization is included. We report on observations obtained in case studies and suggest the above mentioned improvement of the Muller-Schupp algorithm. In Section 4 we comment on the context of our work as well as on perspectives for extensions.

2 The algorithms of Safra and Muller-Schupp

We consider Büchi automata in the format $\mathcal{A} = (Q, \Sigma, q_0, \Delta, F)$ where Q is the finite set of states, Σ the input alphabet, q_0 the initial state, $\Delta \subseteq Q \times \Sigma \times Q$ the transition relation, and $F \subseteq Q$ the set of "final states". The automaton $\mathcal{A}$ accepts the ω-word $\alpha \in \Sigma^\omega$ if a a run $\rho \in Q^\omega$ exists (defined in the standard way) with infinitely many occurrences of states in F. In other words, one may consider the *run tree* t_α of $\mathcal{A}$ on α, which has a root (considered to be at "level 0") labelled q_0, and displays level by level the states reached after the α-prefixes $\alpha(0) \ldots \alpha(i-1)$ for $i = 0, 1, 2, \ldots$. Formally, a vertex on level i labelled p has the successor nodes labelled $q_1, \ldots, q_k$ if for p and the letter $a = \alpha(i)$ the transitions $(p, a, q_1), \ldots, (p, a, q_k)$ are applicable. We assume the tree to be sibling ordered, with reference to an ordering of the set of states. The *run dag* r_α of $\mathcal{A}$ on α is obtained inductively from t_α, level by level, by deleting a vertex v labelled q if on the same level a vertex u labelled q appears to the left; in this case an edge is added from the parent vertex of v to u. Clearly, the input word α is accepted by $\mathcal{A}$ iff in the run dag r_α there is an infinite path from the root on which an F-state occurs infinitely often (henceforth we just speak of a "successful path").

A deterministic Rabin automaton (we say "Rabin automaton" to be short) has the format $\mathcal{A} = (Q, \Sigma, q_0, \delta, \Omega)$ where Q, Σ, q_0 are as for Büchi automata, $\delta : Q \times \Sigma \to Q$ is the transition function, and $\Omega = ((E_1, F_1), \ldots, (E_k, F_k))$ a list of "accepting pairs" with $E_j, F_j \subset Q$. The automaton accepts the input word $\alpha \in \Sigma^\omega$ if for the unique run $\rho \in Q^\omega$ of $\mathcal{A}$ on α an index j exists such that some F_j-state is visited infinitely often in ρ but each E_j-state only finitely often.

The starting point for the transformation of a Büchi automaton $\mathcal{A} = (Q, \Sigma, q_0, \Delta, F)$ into an equivalent Rabin automaton (i.e., recognizing the same ω-language) is to use a finite abstraction of the infinity of the finite prefixes of run trees. If the Büchi automaton has scanned the prefix $\alpha(0) \ldots \alpha(i-1)$ of the input, the run tree up to level i is built up. Taking the run dag instead, one observes that a structure of finite width suffices (since each state can occur at most once on each level). The main point of the transformation of $\mathcal{A}$ into a finite deterministic automaton is to invent a finite number of representations of the infinitely many possible run dag prefixes, in a way that the existence of an infinite path with infinitely many F-states can still be detected. For this, it is necessary to separate the different threads of the run tree (or run dag) for recording of the occurrence of final states. Both algorithms, the procedures by Safra and Muller-Schupp, use tree structures for this purpose. A node in such a tree provides the information which states are presently visited in certain threads of the run tree; in particular, the root records the totality of presently reachable states (as in the classical subset construction). Also both procedures adopt the convention that a state is kept only at its leftmost occurrence on a tree level, thus inheriting the rule mentioned above for constructing run dags.

2.1 Muller-Schupp trees

Let us first introduce the tree structures used by Muller and Schupp, called Muller-Schupp trees in the present paper. A *Muller-Schupp tree* is a finite sibling-ordered strictly binary tree (i.e., each vertex except the leaves has precisely two sons), whose vertices are named with positive natural numbers and additionally are labelled by two items: a subset of Q and a color from the set $\{\text{red}, \text{yellow}, \text{green}\}$. Since by construction the set of states at a parent node is the disjoint union of the sets at the two sons, it would suffice to keep state-sets as labels only for the leaves; however, for easier readability of the trees we prefer to use the state-set labelling throughout.

The Muller-Schupp trees can be motivated in three stages, starting from the computation tree t_α of the given Büchi automaton on some input word α. The first step consists in partitioning the sons of a vertex v into two classes, those which carry a final state and those which carry a non-final state. The former are collected in a set and declared as the label of the left son of v, the latter (non-final) ones form the label of the right son (of course, one of the two sons can vanish). Call the resulting tree with at most binary branching the "acceptance tree" t^1_α. It is easy to verify that

t_α has a successful path iff t^1_α has a path branching left infinitely often

If we keep only the occurrences of a state q which occur leftmost on the respective level of the tree t^1_α, obtaining the tree t^2_α, the equivalence above holds also for t^2_α instead of t^1_α.

The tree t^2_α grows in a deterministic fashion level by level, given α. Note that each level has at most as many entries as there are states in $\mathcal{A}$, so t^2_α is of bounded width. The idea is to take as states of the deterministic Rabin automaton compressed versions of the t^2_α-prefixes, level by level: A path segment from a left son, respectively a right son, or from the root, to the next branching point v is contracted into v. Then a strictly binary tree is obtained. The states of such a path segment are forgotten except some information about the presence of final states, given by three different colors which the remaining vertex v can have: red, yellow, and green. It is clear that the number of such trees is finite.

The update step upon processing an input letter a (corresponding to the extension of t^2_α by one level), is performed by attaching sons to the leaves according to the subset construction, starting from the state set at each leaf. Of course, no son is introduced to a leaf if from none of its states a continuation of the run via a is possible. This case leads to the deletion of the whole path back to the last branching point. Vertex names which are freed by this can be reused, however not in the same update step. In the remaining case a left son, a right son, or both are introduced, depending on whether there are only final states, only non-final states, or both in the resulting state set. When a final state is encountered, the vertex carrying it is colored green. By the cancellation policy (to keep only the leftmost occurrence of a state) and the path compression procedure it can happen that path segments are merged in a single vertex (again setting free name of the spared vertex). In this case the parent vertex may receive final states from a son with which it is merged; we say then that it "receives a new final state".

A vertex is colored red if the path segment it represents has no final state, yellow if it has a final state but did not receive this state in the last step, and it is colored green if it received a final state in the last step, either at a leaf via the subset construction or by a merge step, for example with a vertex previously colored yellow.

Using this update procedure, it turns out that t^2_α has a path branching left infinitely often iff in the sequence of corresponding Muller-Schupp trees some vertex v stays forever from some point onwards and has the color green again and again. This is captured by a Rabin acceptance condition of pairs (E_i, F_i) where i ranges over the finite reservoir of vertex names: E_i contains those trees where i is missing, and F_i has those trees where i occurs colored green.

Formally, the update for a tree t and input letter a is carried out as follows:

Update of Muller-Schupp tree

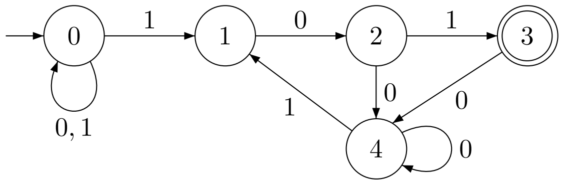
1. Copy the given tree t, changing all colors green to yellow
2. Apply the subset construction (via letter a) to each leaf, add left and right son carrying the reached final, respectively non-final states; color these sons green and red, respectively.
3. Keep only the leftmost occurrence of each state.
4. Delete the vertices which are only on paths leading to leaves whose value is the empty set.
5. As long as there exists a vertex of degree one merge this vertex with its successor, inheriting the color green if this successor was colored green or yellow.
6. Proceeding from the leaves, label each parent by the union of the two state sets from the labelling of the two sons.

2.2 Safra trees

The Safra trees are more succinct in the sense that they suppress as much as possible the record of non-final states. Starting from the update step of the Muller-Schupp algorithm (which of course was not the way these algorithms were invented), the Safra construction introduces a technical simplification when the subset construction is applied: Here only the left son (containing the final states reached) is kept in the tree, no right son for the non-final states is introduced. When from these non-final states at later stages final states are reached, new son vertices are created successively in the Safra tree; in this situation more than binary branching may occur. In a Muller-Schupp tree the intermediate vertices with non-final state-sets amount to a binary encoding of the Safra trees. However, due to different coloring policies, the embedding of a Safra tree into the corresponding Muller-Schupp tree cannot in general be lifted to an embedding of the Safra state space into the Muller-Schupp state space. In particular, a Safra automaton can also be larger than the corresponding Muller-Schupp automaton; cf. the remark at the end of Section 3. A didactic advantage of the Muller-Schupp trees is that they convey more directly the structure of the computation tree of the given Büchi automaton.

The difference of colorings reflects different recordings of visits of final states. The Muller-Schupp procedure uses the coloring policy to signal "new visits to final states". The Safra algorithm marks a node by color green according to the so-called breakpoint construction, signalling that all states of the node can be reached via a *past* visit to a final state. Similarly as for Muller-Schupp trees, a run is accepting if some vertex stays indefinitely from some point onwards and is colored green again and again. For the formal definition of Safra trees and their update function we refer the reader to [6] or [9].

Example 1. Let $\mathcal{A}_0$ be the Büchi automaton

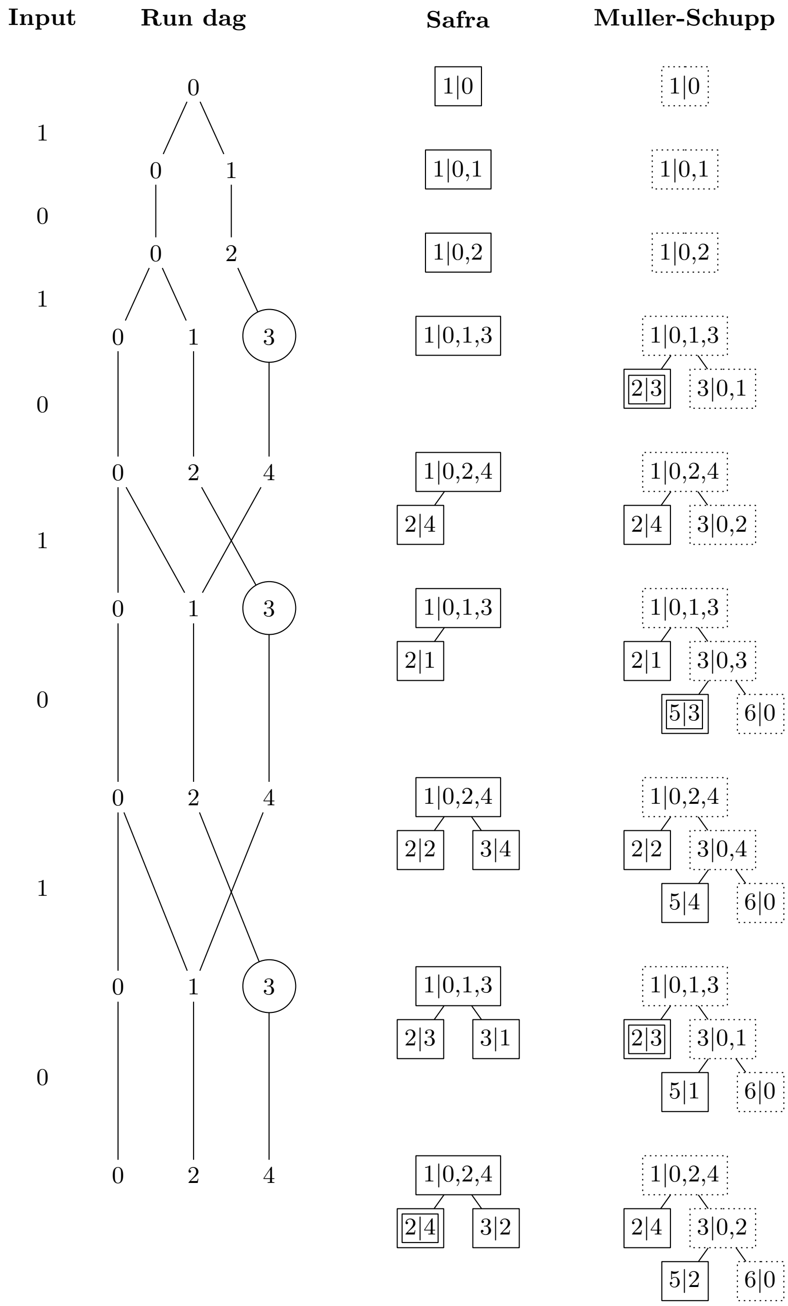


which accepts the 0-1-sequences which have the segment 11 only finitely often but 101 infinitely often. In the figure on the next page we present the run dag for the input word 10101010..., right to it the run of Safra trees, and aside that the run of Muller-Schupp trees. We skip the intermediate stages in the construction process between two successive trees. A vertex is named by a number; following the stroke we list the states belonging to its label. In a Muller-Schupp tree, a vertex colored red is given as a dashed rectangle, colored yellow as a simple rectangle, and colored green as a double-line rectangle. Similarly, the vertices of a Safra tree receiving the mark green are displayed in double-line rectangles.

3 Implementation, experiments, conclusions

OmegaDet is a program written in C++ which offers implementations of four determinization procedures for Büchi automata:

– the Safra construction

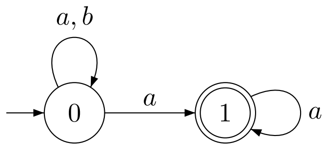

Input
Run dag
Safra
Muller-Schupp

- the Muller-Schupp construction
- an optimized Muller-Schupp construction (which is presented later in this section)
- the Hayashi-Miyano construction (which can be applied to co-Büchi automata)

The fourth option is included since many examples we considered turned out to be co-Büchi automata. This happens, for example, in cases where the accepting loops consist of a single state only. In this case the Büchi condition holds iff from some time onwards only final states are encountered (co-Büchi acceptance). Other examples were not co-Büchi automata as such but could be transformed to this shape by declaring more states (than previously) as final, without changing the accepted language. Thus a useful preprocessing consists in successively declaring as final all states from which in one step only final states are reachable, which does not change the language but might lead to a co-Büchi automaton.

The Hayashi-Miyano construction is a simple refinement of the subset construction, needing only $2^{O(n)}$ states (essentially two subsets) and will be preferable if a large automaton can be presented as a co-Büchi automaton. An implementation exists also in the LASH package at Liège (see [8, 10, 11]).

The program OmegaDet asks the user to supply a Büchi automaton as a text file. The format of text file is best explained by an example. Consider the following Büchi automaton $\mathcal{A}_1$ which accepts all ω-words with only finitely many b.




```
2
ab
1
0 a 0
0 b 0
0 a 1
1 a 1
```

The first line of the text file is the number n of states of the Büchi automaton, whose state set is then assumed to be $\{0, ..., n-1\}$. The initial state is 0. The second line contains the used alphabet Σ as a string of single ASCII symbols. Each symbol of the string is used as a single letter. Afterwards the final states are numerated in the third line separated by spaces. Then the transitions of the automaton are listed.

The user can choose the desired algorithms in a menu. He has the possibility to either simulate a run interactively or to compute the deterministic Rabin automaton. In the latter case, the program reports progress after every 200 computed states. The output is then delivered in four parts:

- the number of states
- the list of Safra trees, respectively Muller Schupp trees, each introduced by a name si, respectively ki (for $i = 0, 1, \ldots$) together with the word (the first in the canonical ordering of words) via which the state is reached, and then a display of the tree (to be explained below),

– the transition table, using the state names si, respectively ki,
– the list of accepting pairs and number of accepting pairs

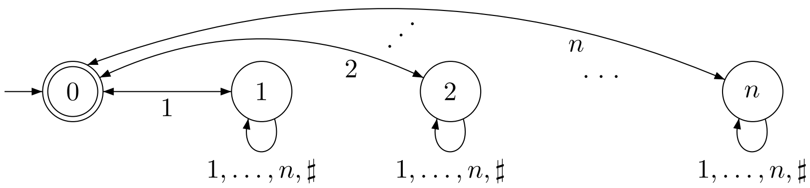
For the display of trees, a textual representation is used which indicates the sons of some node by the subsequent lines, marked as sons by indentation after a pointer symbol `+->`. So brother nodes are listed with same indentation. The colors red, yellow, green are presented as the symbols -, 0, +, following each vertex in Muller-Schupp trees; in Safra trees a mark ! is attached to a vertex if it has color green.

As an example of the output we list the automaton according to Safra generated by the example above.

```
Deterministic Rabin automaton
according to Safra:

4 States:
s0:
    [1|0]

s1: a
    [1|0,1]

s2: aa
    [1|0,1]
     +-> [2|1]

s3: aaa
    [1|0,1]
     +-> [2|1]!
```

```
Transition table:
     a   b
s0  s1  s0
s1  s2  s0
s2  s3  s0
s3  s3  s0

Acceptance pairs:
for vertex 2 (sizes 2,1):
({s0,s1},{s3})

Overall: 1 pair with non-empty
acceptance set
```

We used the program for various Büchi automata, among them also the automata suggested by Michel [4] for showing the $2^{O(n \log n)}$ lower bound for complementation of Büchi automata (see also [3] and [9]). Recall that the Büchi automaton $\mathcal{M}_n$ considered in [4] has states $0, \ldots, n$, the input alphabet $\{1, \ldots, n, \sharp\}$ and the following transition graph:



While we could compute the Rabin automaton according to Safra's construction easily up to automaton $\mathcal{M}_5$ (where a million states are reached for the first time), the Rabin automaton following Muller-Schupp has about 5 million states already for $\mathcal{M}_3$, and the system ran out of memory for $\mathcal{M}_4$ (see Table 1). The reader can obtain an impression of the program OmegaDet by calling `http://www-i7.informatik.rwth-aachen.de/d/research/omegadet.html`; there the program itself, as well as the text files for the automata $\mathcal{M}_1, \ldots, \mathcal{M}_5$ and the output files are listed.

	Safra		Muller-Schupp		Opt. Mulller-Schupp	
	States	Pairs	States	Pairs	States	Pairs
$\mathcal{M}_1$ (2 states)	7	1	9	5	9	5
$\mathcal{M}_2$ (3 states)	33	2	4,058	8	262	7
$\mathcal{M}_3$ (4 states)	385	5	4,823,543	11	23,225	9
$\mathcal{M}_4$ (5 states)	13,601	7	memory exceeded		3,656,802	11
$\mathcal{M}_5$ (6 states)	1,059,057	9	memory exceeded		memory exceeded	

Table 1. Comparison of Safra, Muller-Schupp and optimized Muller-Schupp

It was already mentioned in the previous section that the Muller-Schupp trees tend to be larger due to the inclusion of vertices whose labels are formed from non-final states. A more serious effect, however, is the procedure for introducing new vertices as sons of the leaves and the naming scheme pursued here. Many trees can be generated which have the same structure of vertex labels (and colors), whereas a difference occurs in the vertex naming. An idea to spare vertex names is to add new sons only for leaves which contain final states as well as non-final states. This may result in the freeing of vertex names which may then be reused more quickly.

Formally, we proceed as follows in this modification of the Muller-Schupp algorithm:

Optimized update of Muller-Schupp tree

1. Copy the given tree t, changing all colors green to yellow
2. Apply the subset construction (via letter a) to each leaf.
3. Keep only the leftmost occurrence of each state.
4. For all leaves which contain final states as well as non-final states add left and right son carrying the reached final, respectively non-final states; color these sons green and red, respectively.
5. Color all leaves containing only final states green.
6. Items 4., 5. and 6. of original algorithm

Indeed, this modification can spare many states, as seen in the table above (fourth column). In the example run of the one-page figure above, the optimization is visible in the last four listed Muller-Schupp trees: The vertex names 5,6 are changed there to 4,5. Nevertheless, the number of the Muller-Schupp trees grows still much faster than the number of Safra trees.

More case studies are reported in [12]. A simple general statement relating the numbers of states of the two constructions is not obvious, since we observed several cases where the Safra construction gives a slightly larger automaton than the Muller-Schupp procedure. This happens, for example, for the automaton $\mathcal{A}_1$

considered above, where the Safra construction yields four states and the Muller-Schupp construction (the normal one as well as the optimized one) only two states. The delayed signalization of "success" (by an extended initialization) yields different Safra trees, distinguished by different colorings, for a unique Muller-Schupp tree. A possibility to spare states in the Safra construction, which appears in the exposition of [3], is to exchange the application of the powerset construction and the creation of sons. In the example just mentioned this spares one state, however still leads to a larger Safra automaton than the Muller-Schupp one.

4 Outlook

We have presented the determinization procedures of Safra and Muller-Schupp, working out their similarities and their differences. Remarks on the implementation focussed on the input and output format and the explanation of simple case studies. In Michel's example we found a drastic difference between the two procedures, giving an advantage to Safra's algorithm. We explained this effect and suggested an improvement of the Muller-Schupp algorithm, giving (in some cases) much smaller automata.

As a result of our experimental studies we found a tighter connection between the two algorithms than expected, and we found the superiority of the Safra procedure for building small automata.

Despite much research, the determinization algorithms have so far not reached the stage of application examples of any serious scale, in definite contrast to the many implementations based on Büchi automata and alternating automata, applied mostly in model-checking. On the other hand, determinization is a necessary prerequisite in certain domains, for example in the solution of games with regular winning conditions (where a presentation by nondeterministic automata does not suffice). It seems that before reaching real practice, determinization still needs more investigation, not only with respect to the procedures as such (as done in this paper) but also concerning the phases of "preprocessing" and "postprocessing". The preprocessing phase would involve an analysis of the given Büchi automaton, possibly its reduction e.g. with methods of [13] and the test whether it can as well be represented as a co-Büchi automaton. In the latter case, the simpler determinization procedure of Hayashi and Miyano [8] can be applied (see e.g. [10]). The postprocessing would deal with the reduction or even minimization of deterministic Rabin automata, and it would have to deal with a parameter which we did not consider in the present paper: the complexity of the acceptance condition (number of accepting pairs in a Rabin automaton). So the present paper just addresses a single aspect in the general problem of implementing Büchi automaton determinization.

Acknowledgement

We thank Detlef Kähler for his presentation of [14] in our seminar and remarks on a preliminary version of the paper; we relied on his work at several places. The second author acknowledges fruitful discussions with Christof Löding while preparing a lecture on the Muller-Schupp algorithm.

References

1. McNaughton, R.: Testing and generating infinite sequences by a finite automaton. Information and Control **9** (1966) 521–530
2. Thomas, W.: Automata on Infinite Objects. In van Leeuwen, J., ed.: Handbook of Theoretical Computer Science. Volume B, Formal models and semantics. Elsevier Science Publishers B. V. (1990) 133–191
3. Perrin, D., Pin, J.E.: Infinite Words: Automata, Semigroups, Logic and Games. Volume 141 of Pure and Applied Mathematics. Elsevier (2004)
4. Michel, M.: Complementation is more difficult with automata on infinite words. CNET, Paris (1988)
5. Löding, C.: Optimal bounds for the transformation of omega-automata. In: Proc. of the 19th Conference on Foundations of Software Technology and Theoretical Computer Science. Volume 1738 of Lecture Notes in Computer Science., Springer-Verlag (1999) 97–109
6. Safra, S.: On the Complexity of ω-Automata. In: Proc. of the 29th Symp. on Foundations of Computer Science, Los Alamitos, CA, IEEE Computer Society Press (1988) 319–327
7. Muller, D., Schupp, P.: Simulating Alternating Tree Automata by Nondeterministic Automata: New Results and New Proofs of the Theorems of Rabin, McNaughton and Safra. Theoretical Computer Science **141** (1995) 69–107
8. Miyano, S., Hayashi, T.: Alternating Finite Automata on ω-Words. Theor. Comput. Sci. **32** (1984) 321–330
9. Thomas, W.: Languages, Automata and Logic. In: Handbook of Formal Language Theory. Volume III. Springer-Verlag (1997) 389–455
10. Boigelot, B., Jodogne, S., Wolper, P.: An Effective Decision Procedure for Linear Arithmetic with Integer and Real Variables. ACM Transaction on Computational Logic (to appear) (2005)
11. : (The Liège Automata-based Symbolic Handler (LASH)) Available at http://www.montefiore.ulg.ac.be/∼boigelot/research/lash/.
12. Schulte-Althoff, C.: Construction of deterministic ω-automata: A comparative analysis of the algorithms by safra and muller/schupp. Diplomarbeit, RWTH Aachen (2005)
13. Etessami, K., Wilke, T., Schuller, R.A.: Fair Simulation Relations, Parity Games, and State Space Reduction for Büchi Automata. In: Proc. of ICALP 2001. Volume 2076 of Lecture Notes in Computer Science., Springer-Verlag (2001) 694–707
14. Kähler, D.: Determinisierung von ω-Automaten. Diplomarbeit, Christian-Albrechts-Universität zu Kiel (2001)